A framework for consciousness

Francis Crick and Christof Koch

Here we summarize our present approach to the problem of consciousness. After an introduction outlining our general strategy, we describe what is meant by the term 'framework' and set it out under ten headings. This framework offers a coherent scheme for explaining the neural correlates of (visual) consciousness in terms of competing cellular assemblies. Most of the ideas we favor have been suggested before, but their combination is original. We also outline some general experimental approaches to the problem and, finally, acknowledge some relevant aspects of the brain that have been left out of the proposed framework.

General strategy

The most difficult aspect of consciousness is the so-called 'hard problem' of qualia[1,2]—the redness of red, the painfulness of pain, and so on. No one has produced any plausible explanation as to how the experience of the redness of red could arise from the actions of the brain. It appears fruitless to approach this problem head-on. Instead, we are attempting to find the neural correlate(s) of consciousness (NCC), in the hope that when we can explain the NCC in causal terms, this will make the problem of qualia clearer[3]. In round terms, the NCC is the minimal set of neuronal events that gives rise to a specific aspect of a conscious percept. We discuss elsewhere[4] why we think consciousness has to be largely private. By 'private', we mean that it is accessible exclusively to the owner of the brain; it is impossible for me to convey to you the exact nature of my conscious percept of the color red, though I can convey information about it, such as whether two shades of red appear to me to be the same or different.

Our main interest is not the enabling factors needed for all forms of consciousness, such as the activity of the ascending reticular systems in the brainstem. Rather, we are interested in the general nature of the neural activities that produce each particular aspect of consciousness, such as perceiving the specific color, shape or movement of an object.

As a matter of tactics, we have concentrated on the visual system of primates, leaving on one side some of the more difficult aspects of consciousness, such as emotion and self-consciousness. Our framework may well apply, however, to other sensory modalities. We have been especially interested in the alert macaque monkey because to find the NCC, not only widespread neural activities but also the detailed behavior of single neurons (or small groups of neurons) must be investigated on very fast time scales. This is difficult to do systematically in humans. Methods such as fMRI are too coarse in both space and time to be of much use for our problem. On the other hand, experiments in visual psychology are much easier to do on humans than on monkeys. Moreover, humans can report what they are conscious of and thus convey the 'contents' of their consciousness. For these reasons, experiments on monkeys and humans should be pursued in parallel.

Francis Crick is at the Salk Institute for Biological Studies, 10010 N. Torrey Pines Road, La Jolla, California 92037, USA.
Christof Koch is at the California Institute of Technology, 1200 East California Boulevard, Pasadena, California 91125, USA.
e-mail: koch@klab.caltech.edu

Framework

A framework is not a detailed hypothesis or set of hypotheses; rather, it is a suggested point of view for an attack on a scientific problem, often suggesting testable hypotheses. Biological frameworks differ from frameworks in physics and chemistry because of the nature of evolution. Biological systems do not have rigid laws, as physics has. Evolution produces mechanisms, and often sub-mechanisms, so that there are few 'rules' in biology which do not have occasional exceptions.

A good framework is one that sounds reasonably plausible relative to available scientific data and that turns out to be largely correct. It is unlikely to be correct in all the details. A framework often contains unstated (and often unrecognized) assumptions, but this is unavoidable.

An example from molecular biology might be helpful. The double-helical structure of DNA immediately suggested, in a novel way, the general nature of gene composition, gene replication and gene action. This framework turned out to be broadly correct, but it did not foresee, for example, either introns or RNA editing. And who would have guessed that DNA replication usually starts with the synthesis of a short stretch of RNA, which is then removed and replaced by DNA? The broad framework acted as a guide, but careful experimentation was needed for the true details to be discovered. This lesson is broadly applicable throughout biology.

A preamble on the cerebral cortex

By the 'cortical system', we mean the cerebral cortex plus other regions closely associated with it, such as the thalamus and the claustrum, and probably the basal ganglia, the cerebellum and the many widespread brainstem projection systems.

We shall refer to the 'front' of the cortex and the 'back' of the cortex because the terms 'frontal' and 'prefrontal' can be ambiguous. For example, is the anterior cingulate prefrontal?

The dividing line between front and back is somewhat arbitrary. It roughly coincides with the central sulcus. It may turn out that a good operational definition is that the front is all those parts that receive a significant input, via the thalamus, from the basal ganglia. (This simple division is probably not useful for olfaction, however.)

There is an absolutely astonishing variety and specificity of actions performed by the cortical system. The visual system of the higher mammals handles an almost infinite variety of visual inputs and reacts to them in detail and with remarkable accuracy. It is clear that the system is highly evolved, is likely to be specified epigenetically in considerable detail and can learn a large amount from experience.

The main function of the sensory cortex is to construct and use highly specific feature detectors, such as those for orientation, motion or faces. The features to which any cortical neuron responds

are usually highly specific but multidimensional. That is, one neuron does not respond to a single feature but to a family of related features. Such features are sometimes called the 'receptive field' of that neuron—the 'non-classical receptive field'[5] expresses the relevant context of the 'classical receptive field'. The visual fields of neurons higher in the visual hierarchy are larger and respond to more complex features than those that are lower down.

An important but neglected aspect of the firing of a neuron (or a small group of associated neurons) is its 'projective field'[6]. This term describes the perceptual and behavioral consequences of stimulating such a neuron in an appropriate manner (for further discussion of motor and premotor cortex, see ref. 7). Both the receptive field and the projective field are dynamic, not merely static, and both can be modified by experience.

How are feature detectors formed? A broad answer is that neurons do this by detecting common and significant correlations in their inputs and by altering their synapses (and perhaps other properties) so that they can more easily respond to such inputs. In other words, the brain is very good at detecting apparent causation. Exactly how it does this is more controversial. The main mechanism is probably Hebbian, but Hebb's seminal suggestion needs to be expanded.

All the above might suggest that cortical action is highly local. Nothing could be further from the truth. In the cortex, there is continual and extensive interaction, both among neighboring cells and also very widely, thanks to the many long cortico-cortical and cortico-thalamo-cortical routes. This is much less true of the thalamus itself.

The sensory cortex is arranged in a semi-hierarchical manner. (This is certainly true of the visual cortex[8], but is less clear in the front of the brain.) That is, most cortical areas do not detect simple correlations in the sensory input, but detect correlations among correlations being expressed by other cortical areas. This remarkable feature of the cortex is seldom emphasized.

There has been a great selective advantage in reacting very rapidly, for both predators and prey. For this reason, the best is the enemy of the good. Typically, it is better to achieve a rapid but occasionally imperfect performance instead of a more prolonged one that always produces a perfect result. Another general principle may be to use several rough and ready methods in parallel to reach a conclusion, rather than following just one method very accurately. This appears to be how, for example, people see in depth.

Incoming visual information is often incomplete or ambiguous. If two similar stimuli are presented in rapid succession, the brain blends them together into one percept. If they are different but in contradiction, such as a face and a house, the brain selects one at a time (as in binocular rivalry) instead of blending them together. In cases where there is not enough information to lead to an unambiguous interpretation of one's environment[9], the cortical networks 'fill in'—that is, they make their best guess, given the incomplete information. Such filling-in is likely to happen in many places in the brain. This general principle is an important guide to much of human behavior (as in 'jumping to conclusions').

Our present framework

Having outlined a few general points about the cortical system, let us now consider specifically the NCC and its attendant properties. We are mainly interested in time periods on the order of a few hundred milliseconds, or at the most, several seconds, so we can now leave on one side processes that take more time, such as permanently laying down a new memory. We have listed the main ingredients of our framework under ten headings. We have not previously discussed in print items 3, 5, 7 and 10, though we have mentioned the first three in a book chapter that is still in press[10]. Many of our basic ideas on consciousness, such as the importance of attention and correlated firing, were outlined in our 1990 paper[11], and in 1995, we suggested a plausible function for consciousness[12] and later updated our ideas in a 1998 review[3]. Previously we proposed that people are not directly conscious of the neural activity in primary visual cortex (V1) and that the (visual) cortex appears hierarchical because it contains no strong loops[12,13]. More recently, we have supported the suggestion[14] that in addition to a slower, all-purpose conscious mode, the brain has many 'zombie modes'[15], which are characterized by rapid and somewhat stereotyped responses.

1. The (unconscious?) homunculus

A good way to begin to consider the overall behavior of the cerebral cortex is to imagine that that the front of the brain is 'looking at' the sensory systems, most of which are at the back of the brain. This division of labor does not lead to an infinite regress[16]. (Further discussion of this idea comes at the end of subhead #3 below.)

We have discussed elsewhere[17] whether the neural activity in the front of the brain is largely unconscious. One proposal[18,19], for example, is that humans are not directly conscious of their thoughts, but only of sensory representations of them in their imagination. At the moment, there is no consensus about this[20].

The hypothesis of the homunculus is very much out of fashion these days, but this is, after all, how everyone thinks of themselves. It would be surprising if this overwhelming illusion did not reflect in some way the general organization of the brain.

2. Zombie modes and consciousness

Many actions in response to sensory inputs are rapid, transient, stereotyped and unconscious[14,21]. They could be thought of as cortical reflexes. Consciousness deals more slowly with broader, less stereotyped aspects of the sensory inputs (or a reflection of these in imagery) and takes time to decide on appropriate thoughts and responses. It is needed because otherwise, a vast number of different zombie modes would be required. The conscious system may interfere somewhat with the concurrent zombie system. It seems to be a great evolutionary advantage to have zombie modes that respond rapidly, in a stereotyped manner, together with a slightly slower system that allows time for thinking and planning more complex behavior.

Visual zombie modes in the cortex probably use the dorsal stream in the parietal region[14]. Some parietal activity, however, also affects consciousness by producing attentional effects on the ventral stream, at least under some circumstances. The conscious mode for vision depends largely on the early visual areas (beyond V1) and especially on the ventral stream. There are no recorded cases of purely parietal damage which led to a complete loss of conscious vision.

In a zombie mode, the main flow of information is probably feed-forward. It could be considered a forward-traveling net-wave. A net-wave is a propagating wave of neural activity, but it is not the same as a wave in a continuous medium. Neural networks in a cortex have both short and long connections, so a net-wave may, in some cases, jump over intervening regions. In the conscious mode, it seems likely that the flow is in both directions (see #5 below) so that it resembles more of a standing net-wave[10].

3. Coalitions of neurons

The cortex is a very highly and specifically interconnected neural network. It has many types of excitatory and inhibitory interneurons and acts by forming transient coalitions of neurons, the members of which support one another. 'Coalitions' implies 'assemblies'—an idea which goes back at least to Hebb[22]—plus competition among them (see also ref. 23). On the basis of experimental results in the macaque, some researchers suggest that selective attention biases the competition among rivalrous cell assemblies, but they do not explicitly relate this idea to consciousness[24].

The various neurons in a coalition in some sense support one another, either directly or indirectly, by increasing the activity of their fellow members. The dynamics of coalitions are not simple. In general, at any moment the winning coalition is somewhat sustained, and embodies what we are conscious of.

It may help to make a crude political analogy. The primaries and the early events in an election would correspond roughly to the preliminary unconscious processing. The winning coalition associated with an object or event would correspond to the winning party, which would remain in power for some time and would attempt to influence and control future events. 'Attention' would correspond to the efforts of journalists, pollsters and others to focus on certain issues rather than others, and thus attempt to bias the electorate in their favor. Perhaps those large pyramidal cells in cortical layer 5 that project to the superior colliculus and the thalamus (both involved in attention) would correspond to electoral polls. These progress from early, tentative polls to later, rather more accurate ones as the election approaches. It is unlikely that all this happens in the brain in a fixed time sequence. The brain may resemble more the British system, in which the time between one election and the next can be irregular. Such an analogy should not be pressed too far. Like all analogies, it should be regarded as a possible source of ideas, which of course will have to be confirmed by experiment.

Coalitions can vary both in size and in character. For example, a coalition produced by visual imagination (with one's eyes closed) may be less widespread than a coalition produced by a vivid and sustained visual input from the environment. In particular, representations of imagined visions may fail to reach down to the lower echelons of the visual hierarchy. Coalitions in dreams may be somewhat different from waking ones.

If there are coalitions in the front of the cortex, they may have a somewhat different character from those formed at the back of the cortex. There may be more than one coalition that achieves winning status and hence produces conscious experience. The coalitions in the front may reflect feelings such as happiness and, perhaps, the feeling of 'authorship', which is related to free will[25]. Such feelings may be more diffuse and persist for a longer time than coalitions in the back of cortex. The terms 'affect' and 'valuations' are now being used for what we have traditionally called 'feelings'[18,19]. Our first working assumption (the homunculus) implies that it is better not to regard the back plus the front as one single coalition, but rather as two or more separate coalitions that interact massively, but not in an exactly reciprocal manner.

4. Explicit representations

An explicit representation of a particular aspect of the visual scene implies that a small set of neurons exists that responds as a detector for that feature, without further complex neural processing. A possible probe, or an operational test, for an explicit representation might be whether a single layer of 'neurons' could deliver the correct answer. For example, if a single-layered neural network were fed the activity of retinal neurons, it would not be able to recognize a face. Fed from the relevant parts of inferior temporal cortex, however, it could reliably signal 'face' or 'no face'.

There is much evidence from both humans and monkeys that if there are no such explicit neurons, or if they are all lost by brain damage, then the subject is unable to consciously perceive that aspect directly. Well-known clinical examples are achromatopsia (loss of color perception), prosopagnosia (loss of face recognition) or akinetopsia (loss of motion perception). In all cases, one or a few attributes of conscious experience have been lost, while most other aspects remain intact. In the macaque, a small, irreversible lesion of the motion area MT/V5 leads to a temporal deficit in motion perception that recovers within days. Larger lesions cause a more permanent loss[26].

Note that an explicit representation is a necessary but not sufficient condition for the NCC to occur.

One can describe this in terms of 'essential nodes'[27,28]. The cortical neural networks (at least for perception) can be thought of as having nodes. Each node is needed to express one aspect of one percept or another. An aspect cannot become conscious unless there is an essential node for it. For consciousness, there may be other necessary conditions, such as projecting to the front of the brain[12].

A node, all by itself, cannot produce consciousness. Even if the neurons in that node were firing appropriately, this would produce little effect if their output synapses were inactivated. A node is a node, not a network. Thus a particular coalition is an active network, consisting of the relevant set of interacting nodes that temporarily sustains itself.

Much useful information can be obtained from lesions. In humans, the damaged area is usually fairly large. It is not clear what effects a very small, (possibly bilateral) reversible lesion would have in the macaque, as it is difficult to discover exactly what a monkey is conscious of. The smallest useful node may be a cortical column[29] or, perhaps, a portion of a cortical column. The feature which that node represents is (broadly) its columnar property. This is because although a single type of pyramidal cell usually sends its information to only one or two cortical areas, the pyramidal cells in a column project the columnar property collectively to many cortical areas, and can thus strengthen any coalition that is forming.

5. The higher levels first

For a new visual input, the neural activity first travels rapidly and unconsciously up the visual hierarchy to a high level, possibly in the front of the brain (this might instantiate a zombie mode). Signals then start to move backward down the hierarchy so that the first stages to reach consciousness are at the higher levels (showing the gist of the scene[30,31]; see also ref. 32), which send these 'conscious' signals again to prefrontal cortex, followed by corresponding activity at successive lower levels (to provide the visual details). This is an oversimplified description. There are also many horizontal connections in the hierarchy.

How far up the hierarchy the initial net-wave travels may depend upon whether attention is diffused or focused at some particular level.

6. Driving and modulating connections

In considering the physiology of coalitions, it is especially important to understand the nature of neural connections. The classification of neuronal inputs is still in a primitive state. It is a mistake to think of all excitatory neural connections as the same type. First, connections to a cortical neuron fall roughly into two

Fig. 1. The snapshot hypothesis proposes that the conscious perception of motion is not represented by the change of firing rate of the relevant neurons, but by the (near) constant firing of certain neurons that represent the motion. The figure is an analogy. It shows how a static picture can suggest motion.

broad classes: driving and modulating inputs[13]. For cortical pyramidal cells, driving inputs may largely contact the basal dendrites, whereas modulatory inputs include back-projections (largely to the apical dendrites) or diffuse projections, especially those from the intralaminar nuclei of the thalamus.

This classification may be too simple. In some cases, a single type of input to a neuron may be driving, such as the input from the lateral geniculate nucleus (LGN) to V1. In other cases, several types of 'driving' inputs may be needed to make that neuron fire at a significant rate. It is possible that the connections from the back of the brain to the front are largely driving, whereas the reverse pathways are largely modulatory, but this is not experimentally established. This general pattern would not hold for cross-modal connections.

The above tentative classification is largely for excitatory cells. Strong loops of driving connections probably do not occur under normal conditions[13]. It seems likely that cortical layer 5 cells which project to the thalamus are driving, and that those from layer 6 are modulating[13,33].

7. Snapshots

Has a successful coalition any special characteristics?

We propose that conscious awareness (for vision) is a series of static snapshots, with motion 'painted' on them[34,35] (**Fig. 1**). By this we mean that perception occurs in discrete epochs. It is well established that the mechanisms for position-estimation and for detecting motion are largely separate. (Recall the motion after-effect.) Thus, a particular motion can be represented by a constant rate of firing of the relevant neurons. It is not surprising, then, that if the eyes do not move in smooth pursuit, the brain is very poor at recognizing acceleration even though it is good at distinguishing movements[36]. (As perceived motion is constant during a snapshot, it can only change between snapshots, which suggests that there is little or no explicit representation for such a change. Hence, acceleration is not easily seen). All the other conscious attributes of the percept at that moment are part of the snapshot.

The durations of successive snapshots are unlikely to be constant. (They are difficult to measure directly.) Moreover, the time of a snapshot for shape, say, may not exactly coincide with that for, say, color. It is possible that these durations may be related to the α rhythm[37] or even the δ rhythm. The theory of the 'perceptual moment' was suggested as early as 1955, when it was not known how motion is represented in the brain[38], but in recent years, it has been largely forgotten.

To reach consciousness, some (unspecified) neural activity for that feature has to cross a threshold. It is unlikely to do so unless it is, or is becoming, the member of a successful coalition. It is held above threshold, possibly as a constant value of the activity, for a certain time (the time of that snapshot). As specific attributes of conscious perception are all-or-none, so should be the underlying NCC (for example, firing at either a low or high level). This activity may also show hysteresis; that is, it may stay there longer than its support warrants.

What could be special about this activity that reaches above the consciousness threshold? It might be some particular way of firing, such as a sustained high rate, some sort of synchronized firing or firing in bursts. Or it might be the firing of special types of neurons, such as those pyramidal cells that project to the front of the brain[39] (Fig. 2). This may seem unlikely but, if true, would greatly simplify the problem, both experimentally and theoretically.

What is required to maintain this special activity above the threshold? It might be something special about the internal dynamics of the neuron, perhaps involving the accumulation of chemicals such as Ca^{2+}, either in the neuron itself or in one of its associated inhibitory neurons. It might also be re-entrant circuits in the cortical system[40]. Positive feedback loops could, by iteratively exciting the neuron, push its activity increasingly upward so that the activity not only reaches above some critical threshold, but is maintained there for some time.

There are a few complications: the threshold level may depend on the rate of approach to the threshold or on how long the input is sustained, or both of these. At the beginning of each new snapshot, there may be some hold-over from the previous one.

Put another way, these are partial descriptions of conscious coalitions forming, growing or disappearing.

There is no evidence for a regular clock in the brain on the second or fraction-of-a-second time scale. The duration of any snapshot (or fragment of a snapshot) is likely to vary somewhat, depending on such factors as sudden on-signals, off-signals, competition, habituation and so on. Several psychological effects have been described[41], such as an illusion under constant illumination similar to the wagon-wheel effect, which hint that there are some irregular batch-like effects in vision.

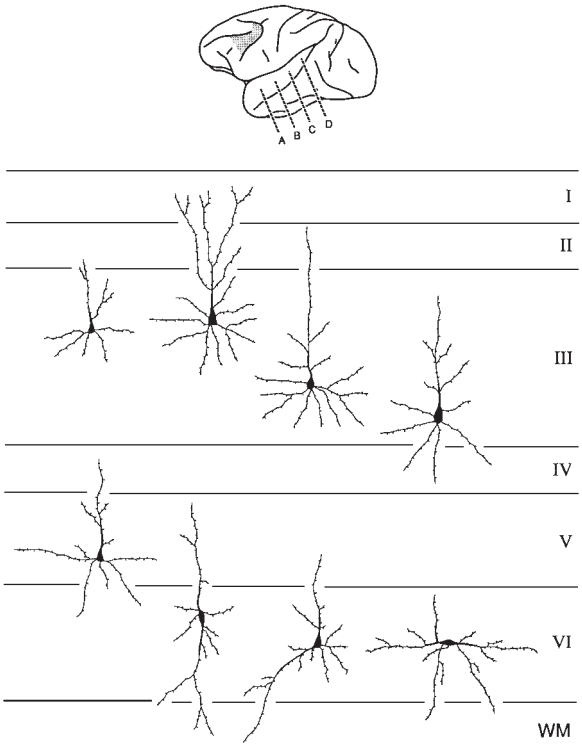

Fig. 2. The dendritic arborization of the different types of neurons in the inferior temporal gyrus of the macaque monkey that project to the prefrontal cortex near the principal sulcus (top, shaded gray). The neurons were recovered from four slices (A–D) as indicated. There are other types of neurons in this area that project to other places. Note that only one type of cell has apical dendrites that reach to layer 1. Drawing from de Lima, A.D., Voigt, T. and Morrison, J.H. (1990)[39], reprinted with permission of Wiley-Liss, Inc., a subsidiary of John Wiley & Sons, Inc.

8. Attention and binding

Attention can usefully be divided into two forms: either rapid, saliency-driven and bottom-up or slower, volitionally controlled and top-down. Each form of attention can also be more diffuse or more focused. Attention probably acts by biasing the competition among rival coalitions, especially during their formation[24]. Bottom-up attention may often start from certain layer 5 neurons that project to parts of the thalamus and the superior colliculus. Top-down attention from the front of the brain may go by somewhat diffuse back-projections to apical dendrites in layers I, II and III, and perhaps also via the intralaminar nuclei of the thalamus (because these have inputs from the front of the brain). Although such projections are widespread, it does not follow that they are nonspecific. To attend to 'red' involves specific connections to many places in cortex. An attractive hypothesis is that the thalamus is largely the organ of attention. The reticular nucleus of the thalamus may help select among attentional signals on a broad scale. Although attention can produce consciousness of a particular object or event by biasing competition among coalitions, activities associated with non-attended objects are quite transient, giving rise to fleeting consciousness (such as the proto-objects suggested in ref. 42).

What is binding? (For reviews that address the binding problem, see *Neuron* 24, 1999.) This is the term used for the process that brings together rather different aspects of an object/event, such as its shape, color, movement and so on. Binding can be of several types[11]. If it has been laid down epigenetically, or has been learnt by experience, it is already embodied in one or more essential nodes so that no special binding mechanism is needed. If the binding required is (relatively) novel, then in some way the activities of separate essential nodes must be made to act together.

Recent psychophysics suggests that 'parallel versus serial' search and 'pre-attentive versus attentive' processing describe two independent dimensions rather than variations along a single dimension (Reddy, L., VanRullen, R. & C.K., *Vision Sci. Soc. 2nd Annu. Mtg.*, abstr. 443, 2002). Their results can all be expressed in terms of the relevant neural networks. Several objects/events can be handled simultaneously—more than one object/event can be attended to at the same time—if there is no significant overlap in any cortical neural network. That is, if two or more objects/events do not have any very active essential nodes in common, they can be consciously perceived. Under such conditions, several largely separate (sensory) coalitions may exist. If there is necessarily such an overlap, then (top-down) attention is needed to select one of them by biasing the competition among them.

This approach largely solves the classical binding problem, which was mainly concerned with how two different objects/events could be 'bound' simultaneously. On this view, the 'binding' of the features of a single object/event is simply the membership in a particular coalition. There is no single cortical area where it all comes together. The effects of that coalition are widely distributed over both the back and the front of the brain. Thus, effectively, they bind by interacting in a diffuse manner.

9. Styles of firing

Synchronized firing (including various oscillations) may increase the effectiveness of a neuron, while not necessarily altering its average firing rate[43]. The extent and significance of synchronized firing in cortex remains controversial[44]. Computations show[45] that this effectiveness is likely to depend on how the correlated input influences the excitatory and inhibitory neurons in the recipient region to which the synchronized neurons project.

We no longer think[11] that synchronized firing, such as the so-called 40 Hz oscillations, is a sufficient condition for the NCC. One likely purpose of synchronized firing is to assist a nascent coalition in its competition with other (nascent) coalitions. If the visual input is simple, such as a single bar in an otherwise empty field, there might not be any significant competition, and synchronized firing may not occur. Similarly, such firing may not be needed once a successful coalition has reached consciousness, when it may be able to maintain itself without the assistance of synchrony, at least for a time[46]. An analogy: after obtaining tenure, you can relax a little.

At any essential node, the earliest spike to arrive may sometimes have the advantage over spikes arriving shortly thereafter[47]. In other words, the exact timing of a spike may influence the competition.

10. Penumbra and meaning

Consider a small set of neurons that fires to, say, some aspect of a face. The experimenter can discover what visual features interest such a set of neurons, but how does the brain know what that firing represents? This is the problem of 'meaning' in its broadest sense.

The NCC at any one time will only directly involve a fraction of all pyramidal cells, but this firing will influence many neurons that are not part of the NCC. These we call the 'penumbra'. The penumbra consists of both synaptic effects and also firing rates. The penumbra is not the result of just the sum of the effects of each essential node separately, but the effects of that NCC as a whole. This penumbra includes past associations of NCC neurons, the expected consequences of the NCC, movements (or at least possible plans for movement) associated with NCC neurons, and so on. For example, a hammer represented in the NCC is likely to influence plans for hammering.

The penumbra, by definition, is not itself conscious, although part of it may become part of the NCC as the NCC shifts. Some of the penumbra neurons may project back to parts of the NCC, or its support, and thus help to support the NCC. The penumbra neurons may be the site of unconscious priming[48].

Related ideas

In the last 15–20 years, there has been an immense flood of books and papers about consciousness. For an extensive bibliography, see T. Metzinger's homepage: www.philosophie.uni-mainz.de/metzinger. Many people have said that consciousness is 'global' or has a 'unity' (whatever that is), but have provided few details about such unity[49]. For many years, Baars[50] has argued that consciousness must be widely distributed.

We are not receptive to physicists trying to apply exotic physics to the brain, about which they seem to know very little, and even less about consciousness.

Dennett[51,52] has written at length about his ideas of "multiple drafts", often using elaborate analogies, but he seems not to believe in the existence of consciousness in the same way as we do. Dennett does consider a limited number of psychological experiments, but neurons, he tells us, "are not my department" (pers. comm.).

Grossberg[53] has written for many years about his "adaptive resonance theory" (ART). This he has developed using very simple neural models. ART involves interactions between the forward and the back pathways, but there is little reference to consciousness.

There are two fairly recent books expressing ideas that overlap considerably with ours. The first of these is by Edelman and Tononi[23]. Their "dynamic core" is very similar to our coalitions. They also divide consciousness into primary consciousness (which is what we are mainly concerned with) and higher-order consciousness (which we have, for the moment, put on one side). They state strongly, however, that they don't think there is a special subset of neurons that alone expresses the NCC.

The second book[54] is by Bachmann, who for many years has used the term "microgenesis" to mean what is happening in the 100–200 ms leading up to consciousness, which is our own main concern as well. He considers carefully the relevant psychological phenomena, such as the different types of masking, but has fewer ideas about the detailed behavior of neurons.

A framework somewhat similar to ours has recently been described by Dehaene and Naccache[55], though they do not elaborate on the snapshot hypothesis or on the neural basis of essential nodes.

General remarks

Almost all of the above ideas have been mentioned previously, either by us or by others. We have given references to some of these. We believe that the framework we have proposed knits all these ideas together, so that for the first time we have a coherent scheme for the NCC in philosophical, psychological and neural terms.

What ties all these suggestions together is the idea of competing coalitions. The illusion of a homunculus inside the head looking at the sensory activities of the brain suggests that the coalition(s) at the back are in some way distinct from the coalition(s) at the front. The two types of coalitions interact extensively, but not exactly reciprocally.

Zombie modes show that not all motor outputs from the cortex are carried out consciously. Consciousness depends on certain coalitions that rest on the properties of very elaborate neural networks. We consider attention to consist of mechanisms that bias the competition among these nascent coalitions.

We suggest that each node in these networks has a characteristic behavior. We speculate that the smallest group of neurons to be worth considering as a node is a cortical column, with its own characteristic behavior (its receptive and projective fields).

The idea of snapshots is a guess at the dynamic properties of the parts of a successful coalition, as coalitions are not static, but constantly changing. The penumbra, on the other hand, is all the neural activity produced by the current NCC, yet not strictly part of it.

We also speculate that the actual NCC may be expressed by only a small set of neurons, in particular those that project from the back of cortex to those parts of the front of cortex that are not purely motor and that receive feedback from there. However, there is much neural activity leading up to and supporting the NCC, so it is important to study this as well as the NCC proper. Moreover, discovering the temporal sequence of such activities (A precedes B) will help us to move from correlation to causation.

The explanation here of this interlocking set of ideas is necessarily abbreviated. We have outlined some of this in more detail elsewhere[10]. A more extended account, together with descriptions of the key experimental data, will be published in a forthcoming book by C.K.[56].

The above framework is a guide to constructing more detailed hypotheses so that they can be tested against already-existing experimental evidence and, above all, to suggest new experiments. The aim is to couch all such explanations in terms of the behavior of identified neurons and in the dynamics of very large neural assemblies.

Future experiments

These fall under several headings. Much further experimental work on small groups of neurons is required for cases in which the percept differs significantly from the sensory input, such as in binocular rivalry[57] and in the many visual illusions[58].

Knowledge of the detailed neuroanatomy of the cerebral cortex needs to be greatly expanded, in particular to characterize the many different types of pyramidal cells in any particular cortical area. What do they look like, where do they project and, eventually, does each have a set of characteristic genetic markers? Are there types of pyramidal cells that do not occur in all cortical areas? When recording spiking activity from a neuron, it would be very desirable to know what type it is, and to where this particular cell projects.

The anatomical methods to characterize types of neurons for the macaque monkey have been available for some years[39] (**Fig. 2**). In the last two decades, very little work has been carried out on cell types and their connectivity, mainly for lack of funds, since such work is not

'hypothesis-driven'. However, because structure is often a clue to function, detailed neuroanatomy is essential background knowledge (of the type the human genome project provides for molecular biology).

On occasion, multi-unit electrodes are chronically implanted into alert patients (for example, to localize seizure onset areas in epileptic patients). With their consent, this can provide sparse but critical data about the behavior of neurons during conscious perception or imagery[59]. It would be very valuable if cortical tissue could be stimulated appropriately with such electrodes to generate specific percepts, thoughts or actions[60].

To study the dynamics of coalitions requires simultaneous recordings on a fast time scale, from small groups of neurons in many places in the brain. This might be done on a primate with a relatively smooth cortex, such as the owl monkey. It would require recording simultaneously from probably a thousand or more electrodes, spaced about 1 mm apart, each capable of picking up spikes from single cells as well as the local field potential.

The immense amount of data this would produce could be displayed visually on a two-dimensional map of the cortical surface, either speeded up or slowed down, so that the eye could grasp the nature of the traveling net-waves as a preliminary to a more detailed study of them. The technical difficulties in recording from so many neurons at once in an alert animal are formidable but not insuperable. To develop the method, one could start with a smaller number of electrodes, more widely spaced, and work on one side of the brain of an animal with the corpus callosum cut.

Omissions

Some major omissions from this discussion are (i) a more detailed consideration of the role of the thalamus (and especially of the intralaminar nuclei), (ii) the actions of the basal ganglia and of the diffuse inputs from the brain stem and (iii) a more detailed scheme for the overall organization of the front of the cortex and for motor outputs. For example, is there some sort of hierarchy in the front? Are there separate streams of information, as there are in the visual system? Do the neurons in the front of cortex show columnar behavior and if so, what for?

On the other hand, our concentration on the NCC, and the postponement of the so-called hard problem of qualia, is part of our strategic approach to the overall problem of consciousness.

Acknowledgments

We thank P.S. Churchland, D. Eagleman, G. Kreiman, N. Logothetis, G. Mitchison, T. Poggio, V. Ramachandran, A. Revonsuo and J. Reynolds for thoughtful comments, O. Crick for the drawing and the J.W. Kieckhefer Foundation, the W.M. Keck Foundation Fund for Discovery in Basic Medical Research at Caltech, the National Institutes of Health, the National Institute of Mental Health and the National Science Foundation for financial support.

RECEIVED 12 SEPTEMBER; ACCEPTED 9 DECEMBER 2002

1. Chalmers, D.J. *The Conscious Mind: in Search of a Fundamental Theory* (Oxford Univ. Press, New York, 1995).
2. Shear, J. *Explaining Consciousness: the Hard Problem* (MIT Press, Cambridge, Massachusetts, 1997).
3. Crick, F.C. & Koch, C. Consciousness and neuroscience. *Cereb. Cortex* **8**, 97–107 (1998).
4. Crick, F.C. & Koch, C. Why neuroscience may be able to explain consciousness. *Sci. Am.* **273**, 84–85 (1995).
5. Allman, J., Miezin, F. & McGuinness, E. Stimulus specific responses from beyond the classical receptive field: neurophysiological mechanisms for local–global comparisons in visual neurons. *Annu. Rev. Neurosci.* **8**, 407–430 (1985).
6. Lehky, S.R. & Sejnowski, T.J. Network model of shape-from-shading: neural function arises from both receptive and projective fields. *Nature* **333**, 452–454 (1988).
7. Graziano, M.S., Taylor, C.S. & Moore, T. Complex movements evoked by microstimulation of precentral cortex. *Neuron* **34**, 841–851 (2002).
8. Felleman, D.J. & Van Essen, D.C. Distributed hierarchical processing in the primate cerebral cortex. *Cereb. Cortex* **1**, 1–47 (1991).
9. Poggio, T., Torre, V. & Koch, C. Computational vision and regularization theory. *Nature* **317**, 314–319 (1985).
10. Crick, F.C. & Koch, C. What are the neural correlates of consciousness? in *Problems in Systems Neuroscience* (eds. van Hemmen, L. & Sejnowski, T.J.) (Oxford Univ. Press, New York, 2003).
11. Crick, F.C. & Koch, C. Towards a neurobiological theory of consciousness. *Sem. Neurosci.* **2**, 263–275 (1990).
12. Crick, F. & Koch, C. Are we aware of neural activity in primary visual cortex? *Nature* **375**, 121–123 (1995).
13. Crick, F. & Koch, C. Constraints on cortical and thalamic projections: the no-strong-loops hypothesis. *Nature* **391**, 245–250 (1998).
14. Milner, D.A. & Goodale, M.A. *The Visual Brain in Action* (Oxford Univ. Press, Oxford, UK, 1995).
15. Koch, C. & Crick, F.C. On the zombie within. *Nature* **411**, 893 (2001).
16. Attneave, F. In defense of homunculi. in *Sensory Communication* (ed. Rosenblith, W.A.) 777–782 (MIT Press and John Wiley, New York, 1961).
17. Crick, F.C. & Koch, C. The unconscious homunculus. in *The Neural Correlates of Consciousness* (ed. Metzinger, T.) 103–110 (MIT Press, Cambridge, Massachusetts, 2000).
18. Jackendoff, R. *Consciousness and the Computational Mind* (MIT Press, Cambridge, Massachusetts, 1987).
19. Jackendoff, R. How language helps us think. *Pragmat. Cogn.* **4**, 1–34 (1996).
20. Crick, F. & Koch, C. The unconscious homunculus. *Neuro-psychoanalysis* **2**, 3–11 and subsequent pages for multi-authored discussion (2000).
21. Rossetti, Y. Implicit short-lived motor representations of space in brain damaged and healthy subjects. *Conscious. Cogn.* **7**, 520–558 (1998).
22. Hebb, D. *The Organization of Behavior: a Neuropsychological Theory* (John Wiley, New York, 1949).
23. Edelman, G.M. & Tononi, G. *A Universe of Consciousness* (Basic Books, New York, 2000).
24. Desimone, R. & Duncan, J. Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.* **18**, 193–222 (1995).
25. Wegner, D. *The Illusion of Conscious Will* (MIT Press, Cambridge, Massachusetts, 2002).
26. Newsome, W.T. & Pare, E.B. A selective impairment of motion perception following lesions of the middle temporal visual area (MT). *J. Neurosci.* **8**, 2201–2211 (1988).
27. Zeki, S.M. Parallel processing, asynchronous perception, and a distributed system of consciousness in vision. *Neuroscientist* **4**, 365–372 (1998).
28. Zeki, S. & Bartels, A. Toward a theory of visual consciousness. *Conscious. Cogn.* **8**, 225–259 (1999).
29. Mountcastle, V.B. *Perceptual Neuroscience* (Harvard Univ. Press, Cambridge, Massachusetts, 1998).
30. Biederman, I. Perceiving real-world scenes. *Science* **177**, 77–80 (1972).
31. Wolfe, J.M. & Bennett, S.C. Preattentive object files: shapeless bundles of basic features. *Vision Res.* **37**, 25–43 (1997).
32. Hochstein, S. & Ahissar, M. View from the top: hierarchies and reverse hierarchies in the visual system. *Neuron* **36**, 791–804 (2002).
33. Sherman, S.M. & Guillery, R. *Exploring the Thalamus* (Academic Press, San Diego, 2001).
34. Zihl, J., Von Cramon, D. & Mai, N. Selective disturbance of movement vision after bilateral brain damage. *Brain* **106**, 313–340 (1983).
35. Hess, R.H., Baker, C.L. Jr. & Zihl, J. The 'motion-blind' patient: low-level spatial and temporal filters. *J. Neurosci.* **9**, 1628–1640 (1989).
36. Simpson, W.A. Temporal summation of visual motion. *Vision Res.* **34**, 2547–2559 (1994).
37. Varela, F.J., Toro, A., John, E.R. & Schwartz, E.L. Perceptual framing and cortical alpha rhythm. *Neuropsychologia* **19**, 675–686 (1981).
38. Stroud, J.M. The fine structure of psychological time. in *Information Theory in Psychology* (ed. Quasten, H.) 174–207 (Free Press, Glencoe, Illinois, 1955).
39. de Lima, A.D., Voigt, T. & Morrison, J.H. Morphology of the cells within the inferior temporal gyrus that project to the prefrontal cortex in the macaque monkey. *J. Comp. Neurol.* **296**, 159–172 (1990).

40. Edelman, G.M. *The Remembered Present: a Biological Theory of Consciousness* (Basic Books, New York, 1989).

41. Purves, D., Paydartar, J.A. & Andrews, T.J. The wagon wheel illusion in movies and reality. *Proc. Natl. Acad. Sci. USA* **93**, 3693–3697 (1996).

42. Rensink, R.A. Seeing, sensing and scrutinizing. *Vision Res.* **40**, 1469–1487 (2000).

43. Singer, W. & Gray, C.M. Visual feature integration and the temporal correlation hypothesis. *Annu. Rev. Neurosci.* **18**, 555–586 (1995).

44. Shadlen, M.N. & Movshon, J.A. Synchrony unbound: a critical evaluation of the temporal binding hypothesis. *Neuron* **24**, 67–77, 111–125 (1999).

45. Salinas, E. & Sejnowski, T.J. Correlated neuronal activity and the flow of neural information. *Nat. Rev. Neurosci.* **2**, 539–550 (2001).

46. Revonsuo, A., Wilenius-Emet, M., Kuusela, J. & Lehto, M. The neural generation of a unified illusion in human vision. *Neuroreport* **8**, 3867–3870 (1997).

47. Van Rullen, R. & Thorpe, S.J. The time course of visual processing: from early perception to decision-making. *J. Cogn. Neurosci.* **13**, 454–461 (2001).

48. Schacter, D.L. Priming and multiple memory systems: perceptual mechanisms of implicit memory. *J. Cogn. Neurosci.* **4**, 255–256 (1992).

49. Bayne, T. & Chalmers, D.J. What is the unity of consciousness? in *The Unity of Consciousness: Binding, Integration, Dissociation* (ed. Cleeremans, A.) (Oxford Univ. Press, Oxford, UK, in press).

50. Baars, B.J. *In the Theater of Consciousness* (Oxford Univ. Press, New York, 1997).

51. Dennett, D.C. *Consciousness Explained* (Little, Brown & Co., Boston, Massachusetts, 1991).

52. Dennett, D. Are we explaining consciousness yet? *Cognition* **79**, 221–237 (2001).

53. Grossberg, S. The attentive brain. *Am. Sci.* **83**, 438–449 (1995).

54. Bachmann, T. *Microgenetic Approach to the Conscious Mind* (John Benjamins, Amsterdam and Philadelphia, 2000).

55. Dehaene, S. & Naccache, L. Towards a cognitive neuroscience of consciousness: basic evidence and a workspace framework. *Cognition* **79**, 1–37 (2001).

56. Koch, C. *The Quest for Consciousness: a Neurobiological Approach* (Roberts and Company Publishers, California, in press).

57. Leopold, D.A. & Logothetis, N.K. Multistable phenomena: changing views in perception. *Trends Cogn. Sci.* **3**, 254–264 (1999).

58. Eagleman, D.M. Visual illusions and neurobiology. *Nat. Rev. Neurosci.* **2**, 920–926 (2001).

59. Kreiman, G., Fried, I. & Koch, C. Single-neuron correlates of subjective vision in the human medial temporal lobe. *Proc. Natl. Acad. Sci. USA* **99**, 8378–8383 (2002).

60. Fried, I., Wilson, C.L., MacDonald, K.A. & Behnke, E.J. Electric current stimulates laughter. *Nature* **391**, 650 (1998).